

中华医学期刊全文数据库

功能概要说明书

一、项目概述

《中华医学期刊全文数据库》是《中华医学杂志》社有限责任公司（后称杂志社或甲方）面向图书馆、医院提供的医学期刊数据库服务平台；旨在探索可持续的数字出版转型道路，提升中华医学会系列杂志的传播影响力及垂直服务能力。

二、项目背景

为巩固百年“中华”牌期刊出版优势、强化医学科学传播能力、搭建医学科研交流主阵地，中华医学会杂志社正逐步全面落实融合出版战略，提振刊群集约化发展，强力推进医学科技期刊的集约化、数字化转型。

《中华医学期刊全文数据库》服务对象分为合作服务商 P，机构用户 B（包括医疗机构、科研院校、医药企业、政府机构），和个人用户 C。主要服务人群是临床医务工作者，其他服务人群包括医学生、图情工作者、政策决策者等。

三、总体目标

近年来，机器学习、自然语言识别、动态出版、基于领域本体的知识库构建

等技术推动了科技期刊数字化转型。科技期刊在生产方式、传播方式、内容服务模式等方面都发生了变革。汲取国内外科技期刊出版平台的相关经验,对数字化生产流程的改造,最根本的措施是将文献内容的结构化工作前置,从源头实现文档XML 数据化,并以唯一的XML 数据在文献的整个生产周期中流转,从而实现出版质量、发布速率、传播能力提升,推进科技期刊规模化和集约化生产水平,促进科技期刊论文大数据中心的建设,提升科技期刊内容运营能力,进而也提升科技期刊国际传播力和影响力。

3.1 建设内容

中华医学期刊全文数据库的建设目的是为机构用户提供 CMAPH 期刊服务平台下所有期刊资源的在线阅读,并通过统一标准化的资源接口服务实现资源的开放、共享,为第三方知识发现系统提供可控、无缝的全文阅读。

3.11 中华医学期刊全文数据库搜索引擎（云服务版）

为机构用户和个人用户提供友好、强大、精准的信息检索服务。目前覆盖的资源包括 CMAPH 期刊平台下的所有论文和图片。

3.12 中华医学期刊全文数据库搜索引擎（镜像服务版）

资源库镜像服务版是为了满足各个数据库厂商及 B 端机构无法通过互联网的方式接入的一种方式,提供将资源库以本地镜像部署方式进行对接服务,通过提供的 SDK 组件 提供给数据库厂商进行集成对接,为 B 端机构提供统一检索、

统一知识发现进行对资源库资源的使用。

3.13 资源库公共服务

资源库云服务版公共服务是为各个数据库厂商及数据库搜索引擎提供一套对外的统一对接服务，可以通过封装的 API 服务进行调用，该服务模式主要是提供资源服务，通过数据库厂商的统一检索、统一知识发现进行对资源库资源的使用。

同时不仅为合作服务商提供题录信息服务，也为合作服务商提供原文的认证和获取服务。

3.14 数据库管理

实现对数据库整体的管理和运营相关服务。

实现对合作服务商用户、机构用户的管理。

3.15 数据中台管理平台

采用微服务架构建立数据中台系统，覆盖全业务场景的需要，全面提升中华医学会数据库的数据治理能力和对外服务归一化能力，能够为各种产品应用提供统一标准化服务。

3.16 知识服务子系统

实现智能推送、实体识别和实体关系查询等功能。

四、系统功能要求

4.1 中华医学期刊全文数据库搜索引擎（云服务版）

搜索引擎是中华医学期刊全文数据库的统一资源检索服务系统。

期刊全文检基于全文检索引擎 Elasticsearch 实现，所有对数据库的配置、监控及操作都通过 Restful 接口完成。数据格式为 json。默认支持节点自动发现，数据自动复制，自动分布扩展，自动负载均衡。适合处理最大千万级别的检索（分布式部署处理效率会大大增加），处理效率非常高。

提供如下功能：

1) 首页

- 申请适用、咨询和反馈
- 首页指南专题库入口
- 首页病例专题库入口
- 首页图标专题库入口
- 用户状态

2) 检索结果处理

- 检索内容清洗
- 智能分词
- 实体识别
- 检索内容智能纠错
- 检索式组织
- 相关度评分

- 最优策略输出

3) 检索结果页输出

- 检索设置
- 检索结果显示
- 在结果中检索
- 分面筛选
- 附带功能区

6) 高级检索页

- 限定检索域
- 文献分类
- 年份限定
- 精确匹配
- 检索表达式组织

7) 历史检索页

8) 专业检索

9) 图表检索

10) 信息聚合页：机构聚合页、作者聚合页、基金聚合页、期刊集合页

11) 帮助说明等说明页

11) 机构管理员管理界面

12) 机构流量管理员管理界面

13) 用户反馈

4.2 资源库公共服务

为知识发现平台提供标准化接口服务和为合作服务商提供数据服务。提供标准化接口和规范，授权机制，服务传输加密等。

4.3 Linkin 共享服务

为了满足第三方平台来源的链入服务，中华医学期刊数据库提供了 Linkin 的服务。

4.4 第三方定制全文检索服务

为了满足对于第三方集成单位对于检索结果页面的集成个性化需求，中华医学期刊数据库系统中有一个基于检索条件，根据检索条件展示检索结果页面的展示，便于具备实现域名 journal.yiigle.com 在中华医学期刊数据库系统的映射和访问，同时将已有的 http://journal.yiigle.com/search_result.zhtml?querys= 功能服务切换为在中华医学期刊数据库系统的对应服务。

4.5 合作商机构管理后台

为合作商提供机构、订单管理和流量统计的平台。合作商通过 CMA 提供的用户名和密码进行登录。登录时限定了合作商的访问 IP，无效 IP 无法登陆系统。

4.6 数据库管理

提供搜索引擎优化、权重设置、特性化词库等功能。

可管理 B 端机构账户，流量机构用户，交易，计费，订单，合作商等。

4.7 网络广告管理

- 1) 广告主维护
- 2) 广告位维护
- 3) 广告物流维护
- 4) 广告订单维护
- 5) 广告投放效果统计

4.8 数据库检测服务包

当 B 端机构用户或 P 端集成商部署系统后，出现无法访问或错误上的问题，可以为 B 端用户提供服务错误提示功能，以便明确发现问题的所在。

4.9 数据中台系统

数据中台系统主要对期刊数据驱动型应用提供数据服务，包括全文检索服务、高级搜索服务、数据分析服务和知识服务，处理一切与资源库和用户行为数据相关操作。

数据中台系统提供四种对外服务，封装为 Restful 接口：

- 基本服务
- 全文检索服务
- 数据分析服务
- 知识服务

需要提供数据治理功能，且容纳机器学习、知识图谱工具和框架。

数据获取模块负责从数据存储层上获得所需数据；数据推送模块负责定时或者动态的按需更新云端数据和内部数据。

4.9.1 基本服务

基本服务提供应用层对数据资源的

- 1) 单一访问功能。
- 2) 批量访问功能。
- 3) 资源列表访问功能。

4.9.2 搜索服务

提供一致性搜索接口，需要比较多个排序算法。

4.9.3 分析服务

需要向服务网关提供数据的各个字段的聚集、过滤等统计分析接口。

4.10 知识服务

需要提供两类接口：智能推荐接口和实体关系查询接口。

知识的构建至少要包含命名实体识别、关系抽取和实体对齐。

提供文档聚类 and 主题模型。

4.11 安全机制

4.11.1 防范体系

防攻击：从服务器、网络、服务漏洞三方面进行考虑设计，服务器操作系统用户权限划分及系统补丁的及时更新，系统杀毒软件保持最新病毒库的更新；网络方面通过硬件防火墙进行非法请求过滤；服务漏洞随时发现随时升级完善。

防盗链 通过防盗链技术的应用可以减轻非正常访问、减轻服务资源的压力，对于全资源进行防盗链技术的实现应用。

防爬虫：爬虫的目的就是大规模地、长时间地获取数据，大规模集中对服务器访问，时间一长就有可能被拒绝。大规模集中访问对服务器的影响较大，爬虫可以短时间增大服务器负载，通过技术手段进行防止网络爬虫。

4.11.2 云安全架构

基于云服务对数据库平台进行安全体系架构，即保障数据库系统高效访问的同时，提升系统的稳定性和安全性。

4.12 系统维护

4.12.1 系统升级更新

系统功能或者服务升级做到不影响、不间断提供对外服务，考虑采用 N 个服务节点，通过部分节点服务进行更新升级，其余进行对外服务，逐批节点进行升级部署。

4.13 资源库镜像服务版

4.14 资源库镜像后台管理

镜像管理后台大致有三类管理行为：账户管理（本地账号系统，与网络账号体系无关）、统计分析、系统管理。

4.15 镜像版资源包升级管理

完成镜像版中所有资源（HTML、图片、PDF）、索引和数据库系统平台的升级管理和实现。

系统非功能要求

软件要求

- 1) 软件开发和运行的基础架构环境应采用 J2EE。
- 2) 数据库管理平台应选用 MySQL。
- 3) 服务器环境使用云服务器。
- 4) 兼容 IE9.0 版本及以上浏览器、Chrome 内核的浏览器、Firefox 浏览器、

Safari 浏览器。

性能需求

1) 并发用户数：在系统并发用户数超过 500 时，系统发生无法连接错误的比例不高于 2%。

2) 检索响应时间：在合理的网络硬件配置情况下，并发用户数不超过 500 情况下，访问操作性界面的系统响应时间不大于 1 秒，页面跳转（如搜索结果跳

转等) 响应时间不超过 3 秒, 在千万级索引记录中查询时间不超过 3 秒。

3) 满足每日 500 万 PV 的访问量。

I/O 处理能力需求

1) 系统提取文档数据的处理过程中, 要求用户前端响应时间 ≤ 10 秒, 文档数据吞吐量过大的可另行设计解决方案。

系统可用性需求

1) 具有良好的管理、监控手段, 可对系统各模块、操作系统、数据库及应用等进行管理监控, 除具备有限的自恢复功能外, 还可采用多种方式进行报警通知管理员。

2) 具备容错手段, 允许操作人员有限范围的误操作和返回。

3) 提供友好、方便的功能界面。支持通用的浏览器阅读。

4) 提供完善的系统操作日志以及异常日志方便故障后进行分析提供依据。

5) 提供系统出现故障时, 给管理员发送提醒(邮件、短信等)。

系统可靠性需求

1) 系统保证 7*24 小时不间断工作。系统无单点故障。

2) 系统具有对硬件、软件运行状态的远程监控和管理能力。

3) 应用软件应有容错能力, 软件故障不引起各类严重的系统再启动。

售后服务需求

产品保修期(维护期)

所采购的软件产品维护期为 1 年(产品维护期从通过验收之日算起)。在产

品维护期内，乙方提供免费的产品维护、版本升级及技术支持服务，对于产品的安全问题应及时提供免费补丁。

服务标准及内容

热线电话服务

此服务终身提供。

- (1) 技术支持工程师直接同客户对话，帮助解决客户提出的疑难问题
- (2) 根据问题的严重程度，将优先解决客户认为是关键而紧急的任务
- (3) 对客户提出的一般性问题进行技术咨询、指导
- (4) 电话热线提供 7×24 的服务能力。即每周 7 天，每日 24 小时提供服务

务

现场技术支持

根据问题的严重程度及用户的要求，在 8 小时内派遣技术支持工程师到用户现场帮助用户解决疑难问题。

此服务 1 年维护期内免费提供。

电子邮件热线服务

可以以电子邮件的形式随时向公司电子邮件热线发送关于各类产品的问题请求，公司专门的技术支持工程师实时监控电子邮件信息，并及时予以回复。

此服务终身提供。

软件产品的更新及维护

此服务 1 年维护期内免费提供。

- (1) 提供具有新增功能的更新软件产品

(2) 提供主要维护版本的更新软件产品

培训计划

要求具备多样化培训方式，避免“讲完就算，讲完就走”的呆板的培训方式，走“减少讲授、提倡研讨、丰富实习”的多角度培训趋势，让系统使用和维护人员在最短的时间内快速、准确地掌握系统的操作运行方式和内核思想。

项目成果

本项目成果以电子文档或者软件形式交付：

(1) 设计分析文档

系统需求分析文档。

系统概要设计文档。

系统详细设计文档。

系统测试文档。

(2) 项目管理文档

系统方案说明书。

项目开发计划。

项目进度管理文档。

系统开发总结报告。

(3) 完整的系统安装程序

(4) 系统安装手册

本系统服务器的安装说明。

本系统客户端需要进行的设置说明。

(5) 系统运行手册

(6) 用户使用手册

即详细的用户操作指南，必须配系统中截图的方式进行编写。

系统帮助文档中的在线帮助表达成果。

(7) 项目管理文档

项目管理各阶段的质量、监控纪录。

项目成果检查纪录。

注：1、版权归甲方共有；2、投标人提供的所有软件产品，要保证是合法的商品软件，若由此引起的第三方知识产权纠纷由投标人负责。

其他需求

1) 该系统涉及到的软件涉及招标方所有期刊的相关文献数据时，文献数据由招标方提供接口提供。所有服务通过接口方式获得。

2) 该系统上所有资源的版权和使用权均归属于招标方。

3) 双方需要签署必要的保密协议。

4) 甲方在该项目投入研发人员以协同乙方在核心项目中的开发，为此乙方需安排好甲方研发人员的研发任务。

5) 在本系统的实施开发实施中，要求成交投标人的系统开发技术人员到甲方做必要的前期开发需求调研，在需求调研的基础上编写需求分析报告和总体技术方案。

6) 提供有丰富的行业经验和开发技能的项目小组进行全过程的实施开发工作，项目组必须熟悉国土管理业务，核心成员必须在实施过同类项目。

7) 具有快速实施能力，能够在要求的时间能完成系统建设目标。

8) 整个项目的实施分为项目准备、项目实施、试运行、项目验收、售后服

务和技术支持几个阶段，针对定制应用软件系统开发，必须完成下述过程：用户需求项目的开发前调研；需求分析；系统设计；详细设计；编码；调试；系统集成及试运行；测试与验收；交付及合同期内的维护。

9) 系统必须提供详细的文档：包括用户手册、管理员维护手册、需求分析报告、系统设计方案、数据库设计说明书等相关开发文档。